周波数スペクトルの概形に着目した母音生成手法

荒川　正和*　林　晃平**　中山　翼***　丸山　晃生*

Vowel generation method focused on outline of frequency spectrum

Masakazu ARAKAWA*, Kouhei HAYASHI**, Tsubasa NAKAYAMA***, and Akio MARUYAMA*

We propose a vowel generation method focused on outline of frequency spectrum for a vowel. One of some vowels is sampled, and modifying it according to the feature of frequency spectrum for each vowel. In this way, the feature of tone is kept and other four vowels are generated from one vowel. We have experimented the case of other four vowels were generated from each of vowels /a/, /i/, /u/, /e/, /o/, and evaluated for five vowels obtained by the above method by four test subjects. We obtained good results in the experiment that vowels of /a/, /i/, /u/, and /o/ are generated by modified vowel /e/ .

Keywords : ***vowel sound, voice synthesis, formant frequency, frequency spectrum, band elimination filter***

1．はじめに

近年，ヤマハ株式会社の VOCALOID™ や，飴屋 / 菖蒲がフリーウェアとして開発した歌声合成ツール UTAU など，音声合成技術が目覚ましい発展を遂げ，注目を集めている．しかしこれらの技術では，歌声の元となるサンプルボイスのライブラリーが大量に必要である．そのため，ソフトウェア自体の容量が大きくなったり，その音声を用意するために多大な時間を要するなどの問題点がある．

そこで，これらの問題を解決するために，ある人の母音を一つサンプリングするだけで，声色の特徴を保ったまま他の母音を生成できるような機構について考える．

本報告では，まず一般的に知られている母音発生の仕組みについて述べ，その考察をふまえ実際に 10 ～ 20 歳代の複数人からサンプリングした母音音声の周波数スペクトルを調べて，母音毎に固有の特徴について考察する．その結果に基づき，サンプリングした音声の声色情報（に影響する周波数成分）を保持したまま，一つの母音をフィルタリングし，別の母音の生成を試みた結果について述べる．最後に，この手法によって得られた母音の質について考察し，今後の課題について述べる．

2．発声の機構

音声が生成されるためには，何らかの駆動源（音源）が必要であり，母音の場合にはその音源が声帯で作られる．声帯で最初に発生される音は倍音のない正弦波であり，それが声道や鼻腔，口腔（歯，口，舌等）の形によって反響，増幅されて周波数特性が変化し，様々な倍音成分を持った声となる．図 1 に，人間の発声器官の構造を示す．

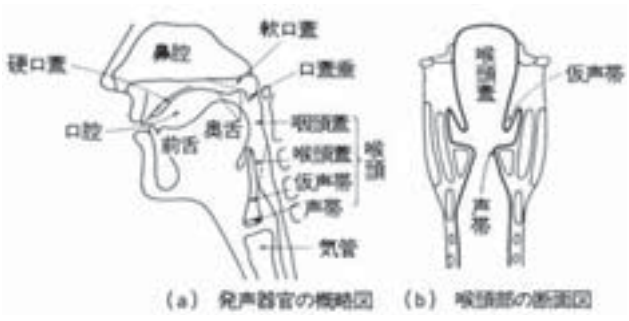
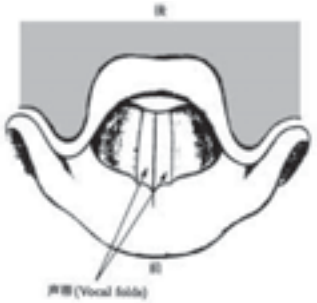


図 1．発声器官の構造[1)]

*電気電子工学科　**㈱アートテクノロジー　***電気電子工学科 5 年

肺から押し出される呼気は，図１(a) に示すように気管の上の喉頭にある３つの弁を通過する．これら３つの弁のうち，一番下についているものが声帯であり，これが振動して音源となる．他の２つは，それぞれ喉頭蓋，仮声帯と呼ばれ，前者は呼気を完全に止める働きを持つ．音源となる声帯の概形と，その振動サイクルを図２に示す．声帯下の圧力は肺からの空気流量が増えるにつれて上昇し，ついには声帯の弁の間を押し広げる．すると，空気が弁の間の空間を通って声道に流れ込む．いったん空気が流れる



(a)

(b)

図２．声帯の構造と振動のサイクル[3)]

と弁下の圧力は急減し，吸引力（ベルヌーイ力）が生じて弁の持つ弾性力も加わって，２つの弁が再び接近する．以上のようなサイクルで空気を断続的に流すことによって，音エネルギーの元（音源）を発生させている．さらにこのとき，声帯筋を緊張させると，この弁に張力が加わるために弁の開閉サイクルが速くなり，声が高くなる．すなわち，声帯の振動周期がピッチ（声の高さ）を決定する．

次に，各母音の生成過程について述べる．音波が円筒管（音響管）を通過すると，管の形状によって決まるある特定の周波数を持つ音波が強められ，別のある周波数は弱められるという共鳴現象が生じる．声帯から唇までの声道を１つの音響管とみなすと，円筒管の場合と同様に共鳴現象が生じることがわかる．

声道の長さは成人でおよそ 17 ㎝程度であり，声道を 17 ㎝の長さを持つ円筒管とみなすと，図３に示すような共鳴現象が生じる．

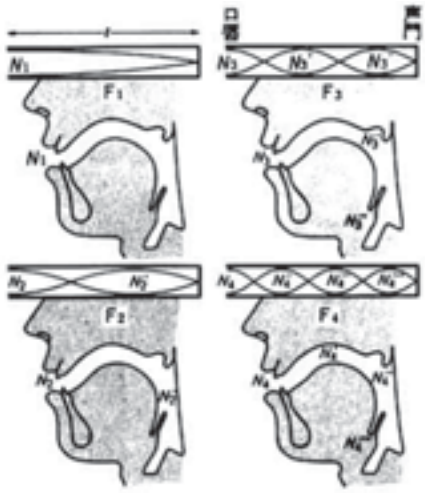


図３．声道における共鳴[1)]

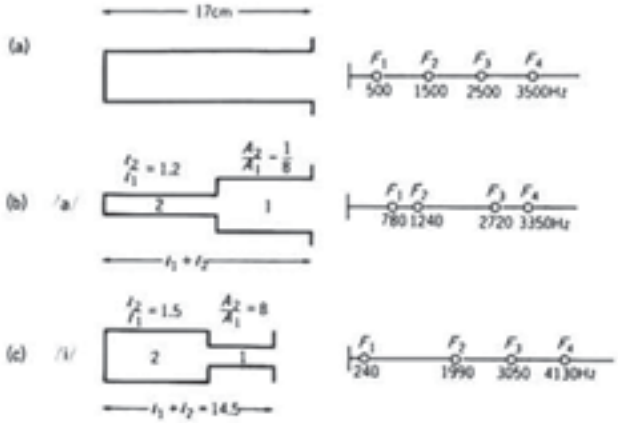


図４．声道の形とホルマントの関係[1)]

実際の声道は，母音によってより複雑な形状を示すが，その形状があまり変化しない部分と大きく変化する部分に分け，単純化して考える．すると図４に示すように，長さと断面積がそれぞれ異なる管を接続した構造のように声道をみなすことができ，声道の形と母音毎の周波数スペクトルのピーク（ホルマント）との間にある，おおよその関係を知ることができる．たとえば，図４(b) は母音の /a/ に相当する声道の形を示している．このとき，第１番目と第２番目の周波数スペクトルのピーク，すなわち第１ホルマント，第２ホルマントはそれぞれ約 780Hz，約 1240Hz となる．図４(c) は，母音 /i/ に相当する声道の形を示しており，第１ホルマント，第２ホルマントはそれぞれ約 240Hz，約 1990Hz となる．

ここで述べた声道の形は，実際には舌や唇を動かすことによって実現されている．図５に，舌の位置とホルマントとの関係を示す．口腔の横断面を見ると，舌の最も盛り上がっているところがあり，これを‘舌の位置’と呼ぶ．この位置と母音の種類には密接な対応関係があり，舌の位置の推移を線で結び母音との関係を示すと，図５(a) に示したような形になる．この舌の位置と第１，第２ホルマント周波数の間には，図５(b) に示すような対応関係があることが知られている．

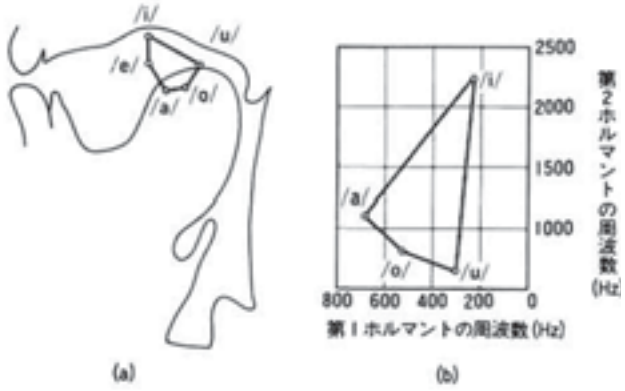


図５．舌の位置とホルマントの関係[1)]

３．母音スペクトルの解析

前節で述べた発声機構をふまえると，母音毎の特徴量としてホルマントがあり，同一の母音であれば異なる人物が発音した音声であってもその周波数スペクトルには共通する特徴がみられるのではないかと考えられる．そこで，異なる人物の同じ母音，および同じ人物の異なる母音の周波数スペクトルを観察・比較し，それぞれに共通する特徴を調べる．各人の母音 /a/，/i/，/u/，/e/，/o/ の音声について，同一母音で共通の特徴，異なる母音間での異なる特徴が見出せれば，それらが母音生成の手掛かりになり得ると予想される．

異なる人物の同じ母音（ここでは /a/）の４種類の周波数スペクトルを図６に示す．それぞれ，(a) 二十代男性A，(b) 二十代男性B，(c) 十代女性[4)]，(d) 十代男性の音声をサンプリングし，PC により周波数スペクトルを解析した．この結果から，まずは周波数スペクトルのピークに着目し，母音毎の特徴を調べることを試みた．

スペクトル解析のそれぞれの周波数とその振幅を数値データとして書き出し，振幅の大きい順に周波数スペクトルをソートし，特徴の抽出を試みる．このとき，これらを単純に振幅順にソートすると各スペクトルピークの裾にある値まで含まれてしまい，適切な結果が得られない．この問題を解決するため，振幅を周波数で微分し，その結果が正から負に変わる点を周波数スペクトルのピークとして採用する．

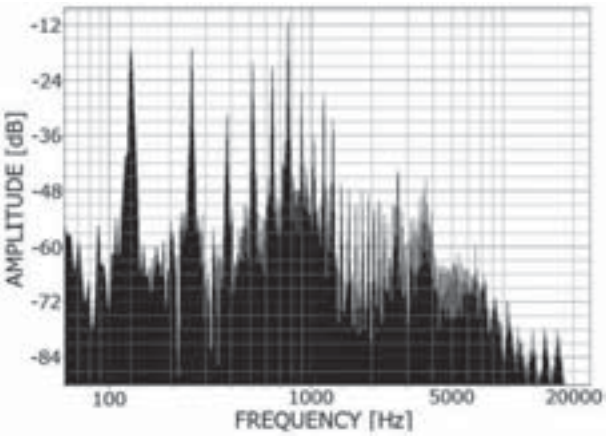
振幅を $I(\upsilon)$，周波数を υ とすると，微分の定義より

$$\frac{\Delta I(\upsilon)}{\Delta \upsilon} = \frac{\Delta I(\upsilon+\Delta\upsilon) - I(\upsilon)}{\Delta \upsilon}$$

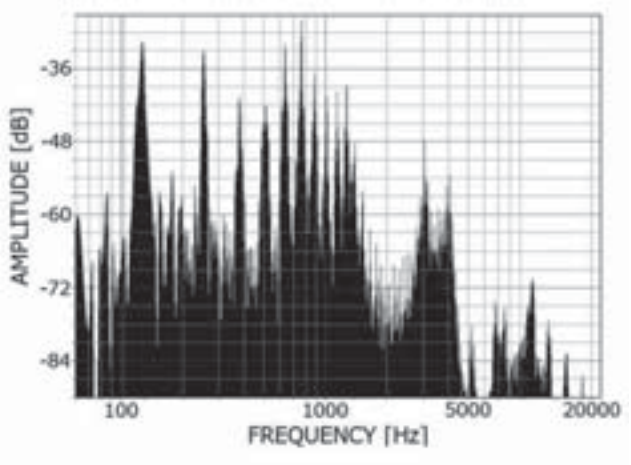
となる．この式を用いて周波数スペクトルのピークを選出し，選出されたピークを振幅順にソートする．ここでは，対象としている音声の音程がそれぞれ異なるので，基音と倍音の比を求め，それらを比較することとした．

それぞれの周波数スペクトルの上位 10 個のピークの周波数と振幅，および倍音の関係を表１に示す．この結果より，どの音声についても１～６倍音が上位のピークとして現れていることが確認できる．しかし，その順番については，有用な規則性を見出すことは出来なかった．このことから，ピーク値の大きな周波数スペクトルのみに注目しても，母音を特定することは困難であることがわかった．

そこで，図６のグラフ全体を見て，その概形に注目してみる．すると，(a) ～ (d) の全てのスペクトルについて，1400 ～ 3000[Hz] にピークの谷があるという共通点が見つかった．この特徴は母音 /a/ に固有のものなのか，それとも人の声固有のものであるかを確認するため，さらに /a/ 以外の母音についても周波数スペクトルの比較を行った．図７～ 10 に，図６と同じ被験者からサンプリングした母音 /i/，/u/，/e/，/o/ の周波数スペクトルの解析結果を順に示す．

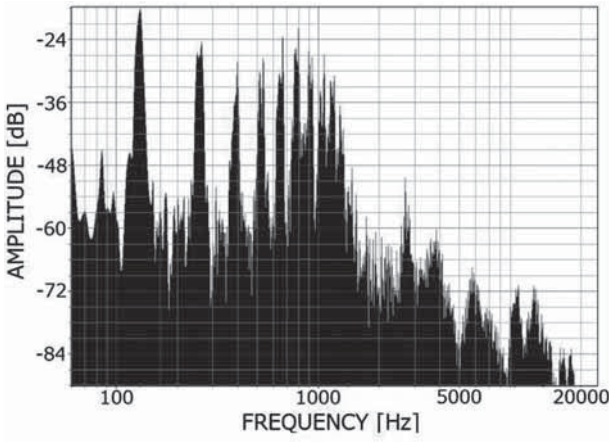


（a）二十代男性Ａ

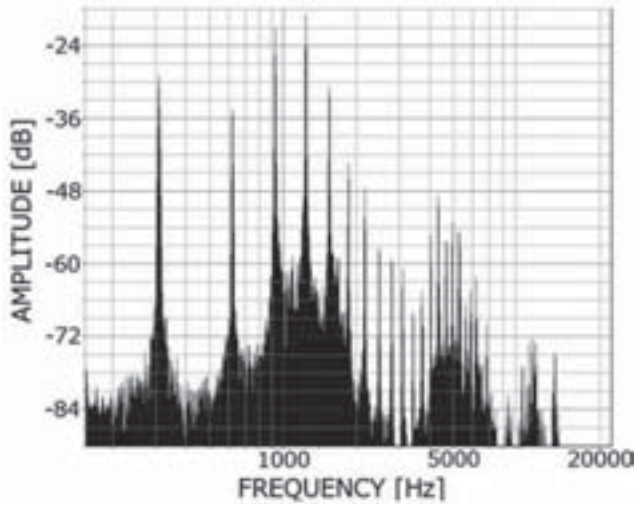


（d）十代男性

図６．/a/ の周波数スペクトル



（b）二十代男性Ｂ



（c）十代女性

表１．倍音成分の比較

（a）二十代男性Ａ

周波数（Hz）	レベル（dB）	倍音
767.1204	− 9.4955	6.08
255.7068	− 17.8052	2.02
759.0454	− 18.5223	6.02
126.5076	− 19.0142	1.00
511.4136	− 20.0019	4.05
640.6128	− 21.0979	5.08
1152.026	− 26.0331	9.14
896.3196	− 26.6023	7.11
1143.951	− 28.4951	9.07
1278.534	− 31.843	10.1

（b）二十代男性Ｂ

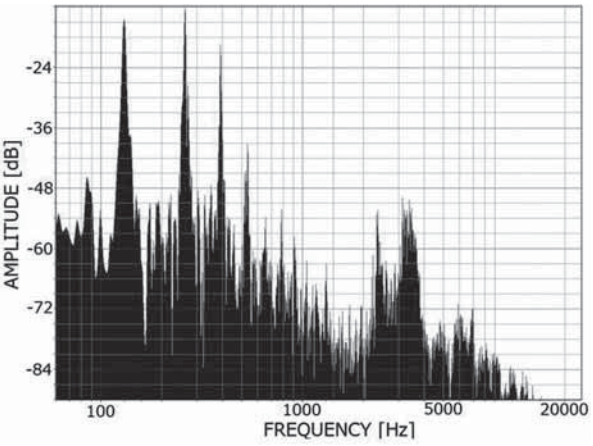
周波数（Hz）	レベル（dB）	倍音
123.0469	− 17.8426	1.00
744.1406	− 21.6218	6.04
729.4922	− 23.9428	5.93
246.0938	− 24.039	2.00
621.0938	− 24.324	5.04
867.1875	− 25.2229	7.05
852.5391	− 25.5232	6.93
609.375	− 27.1808	4.95
369.1406	− 27.6606	3.00
489.2578	− 27.6755	3.97

(c) 十代女性

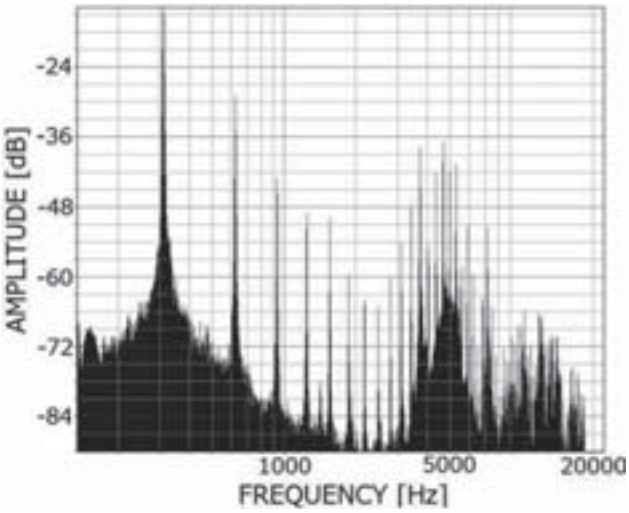
周波数 (Hz)	レベル (dB)	倍音
1240.851	− 18.2757	4.02
1232.776	− 19.6386	3.99
928.6194	− 20.088	3.01
309.5398	− 28.6777	1.00
1550.391	− 28.8682	5.02
1542.316	− 30.2774	4.99
619.0796	− 32.3173	2.00
1859.93	− 41.5347	6.02
1849.164	− 42.8331	5.98
4330.865	− 44.2743	14.0

(d) 十代男性

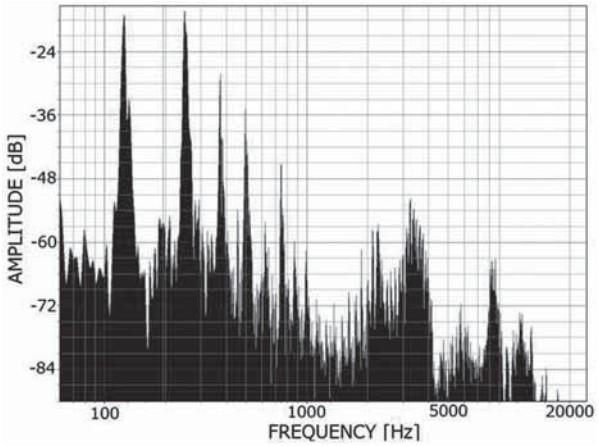
周波数 (Hz)	レベル (dB)	倍音
495.263672	− 28.393564	4.03
255.706787	− 28.778706	2.08
514.105225	− 29.044132	4.18
126.507568	− 30.812952	1.03
759.04541	− 32.058407	6.17
503.338623	− 34.090015	4.09
632.537842	− 35.717949	5.14
376.831055	− 37.109516	3.06
395.672607	− 38.723904	3.22
788.653564	− 39.527489	6.41



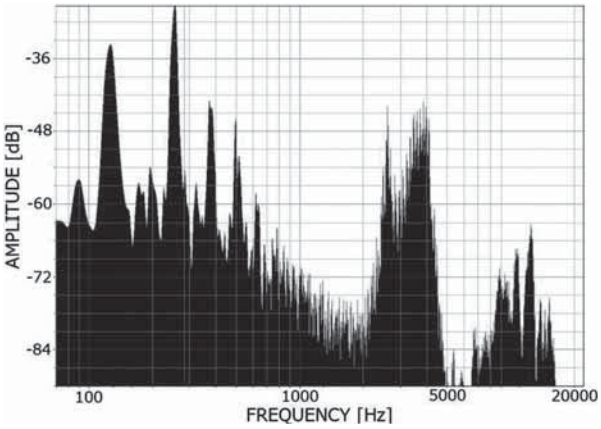
(b) 二十代男性B



(c) 十代女性

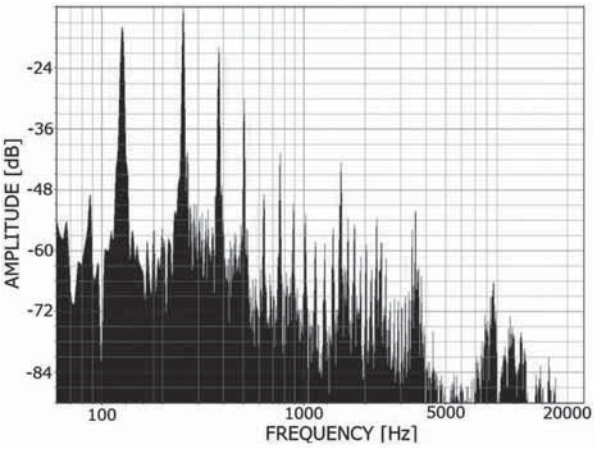


(a) 二十代男性A

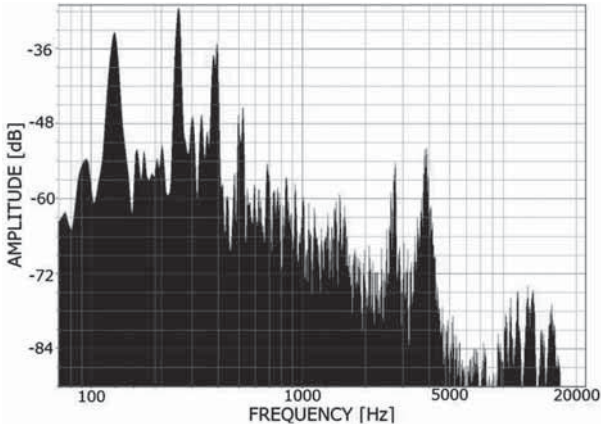


(d) 十代男性

図7．/i/の周波数スペクトル

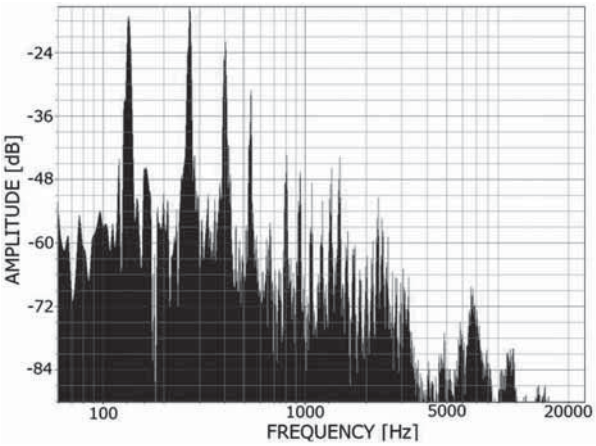


（a）二十代男性Ａ

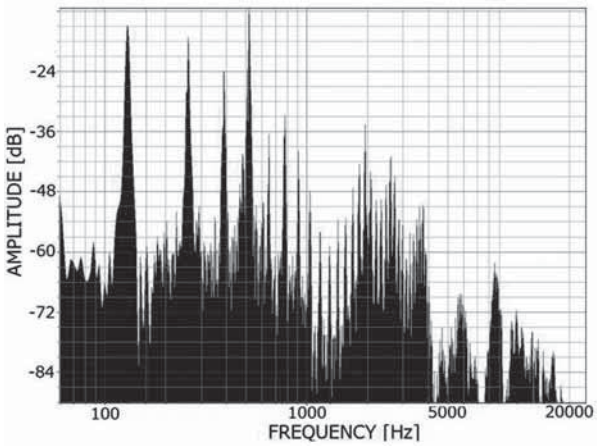


（d）十代男性

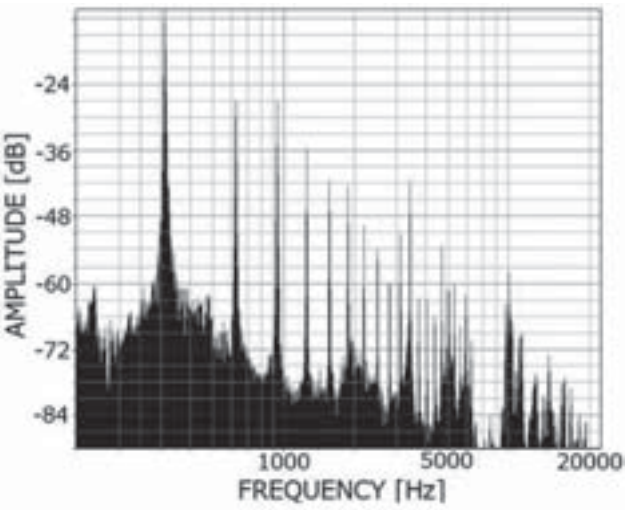
図８．/u/ の周波数スペクトル



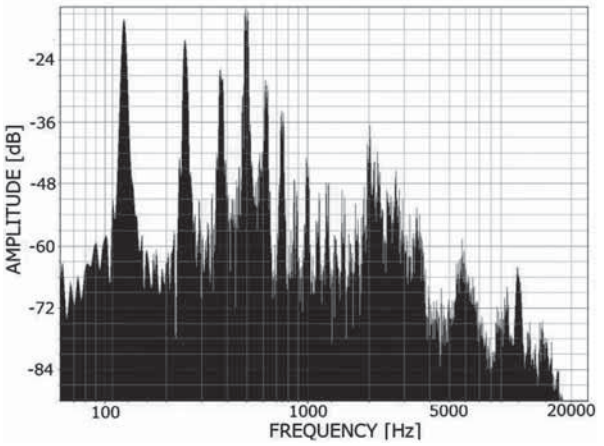
（b）二十代男性Ｂ



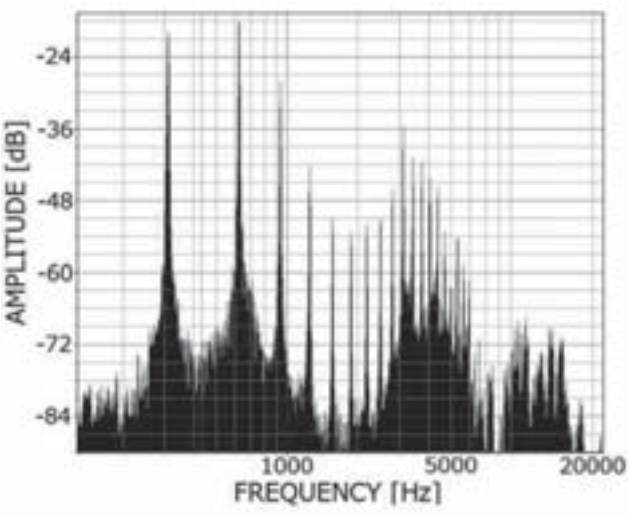
（a）二十代男性Ａ



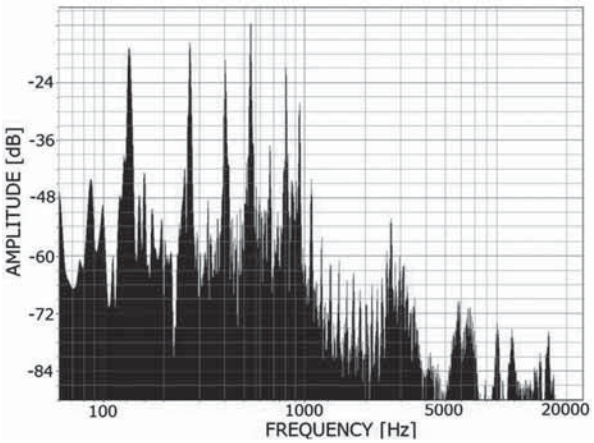
（c）十代女性



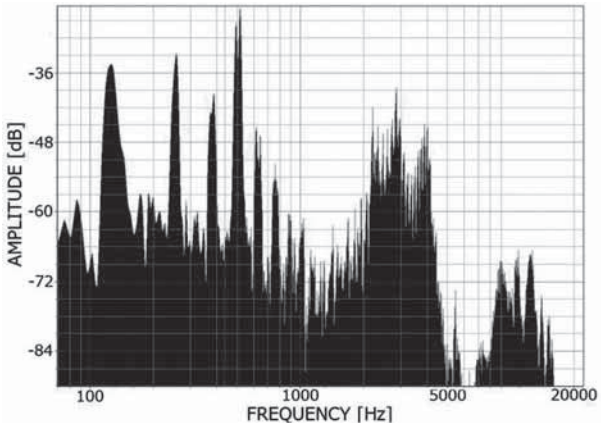
（b）二十代男性Ｂ



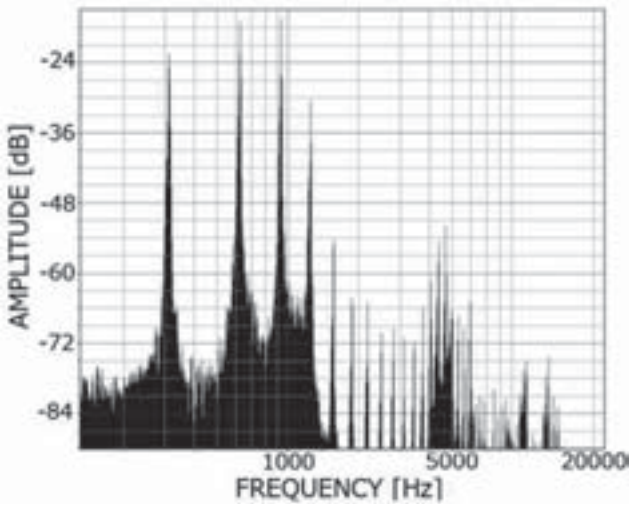
（c）十代女性



（b）二十代男性Ｂ

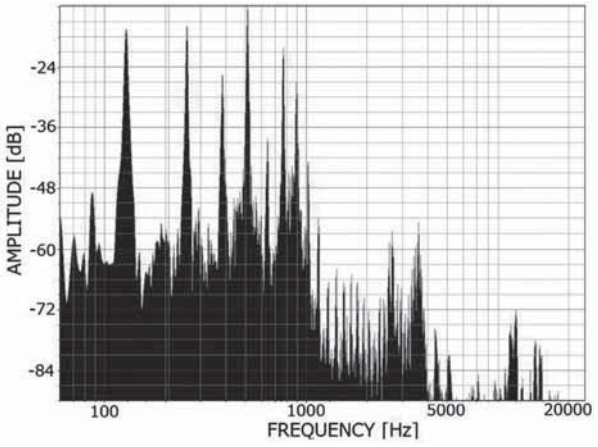


（d）十代男性

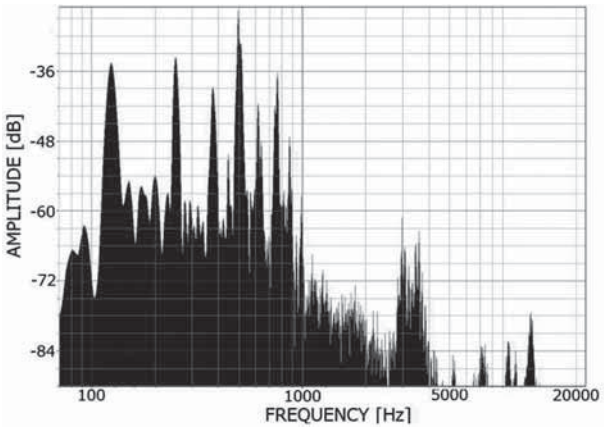


（c）十代女性

図９．/e/ の周波数スペクトル



（a）二十代男性Ａ



（d）十代男性

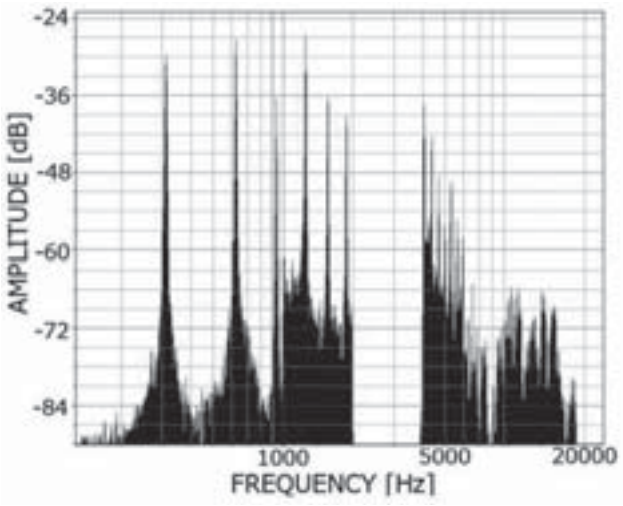
図 10. /o/ の周波数スペクトル

図 7 ～ 10 を比較すると，異なる母音はスペクトルの概形に明らかな相違点があり，同一の母音は似たスペクトルの概形（傾向）を示していることがわかる．

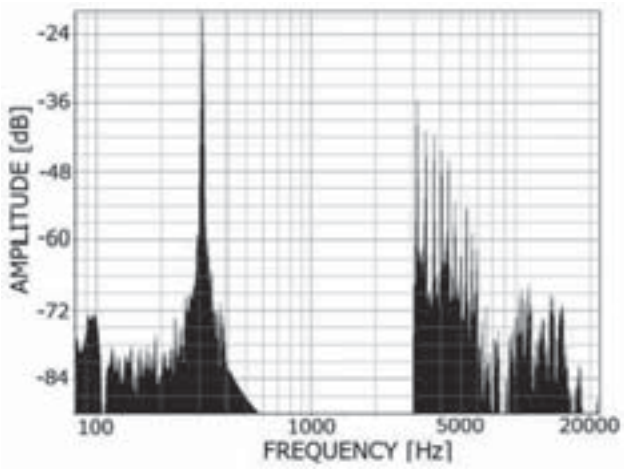
具体的には，/a/ は 1000 ～ 2000Hz の振幅が大きく，2000 ～ 3000Hz の振幅が低い．/i/ は 300Hz ～ 3000Hz 付近まで振幅が下がり続け，以降大きくなる．/u/ は基音が最も高く，そこから上の周波数で振幅が下がり続けている．/e/ は 1000 ～ 3000Hz の振幅が低く，以降大きくなる．/o/ は 1400 ～ 4000Hz の振幅が低く，4000Hz 以上で幅が大きくなっている．以上述べた母音毎の特徴に基づいて，一つの母音から他の 4 母音の生成を試みる．

4．母音の生成と評価

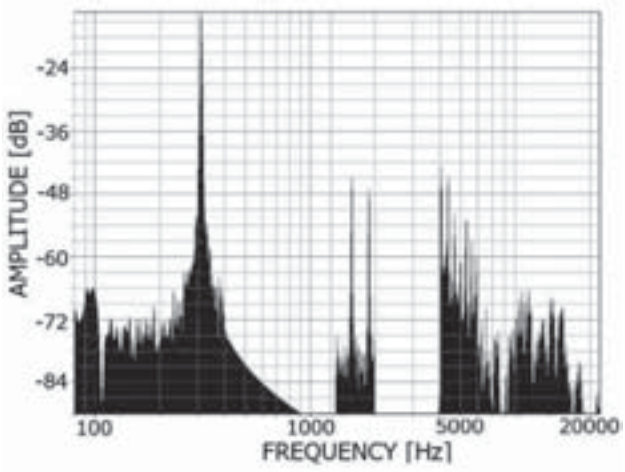
最も多くの周波数成分を含んでいる /e/ 音を用い，前節で着目した各母音の周波数スペクトルの特徴に合わせて PC 上のソフトウェアイコライザで /e/ 音の周波数スペクトルを加工（フィルタリング）して，他の母音 /a/，/i/，/u/，/o/ の生成を試みた．その一例として，十代女性の /e/ 音（図 9 (c)）を加工し，他の 4 母音を生成した場合の周波数スペクトルを，図 11 に示す．これらの周波数スペクトルを持つ音声を 12 人（10 代 3 人，20 代 7 人，40 代 2 人）の被験者に聞かせ，5 種類の母音として認識できるかどうか判定してもらったところ，12 人全員が 5 種類全ての母音 /a/ ～ /o/ について，正しく認識できた．



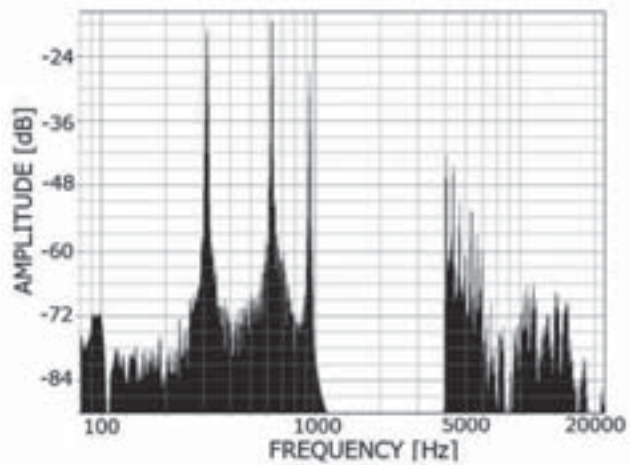
(a) /a/



(b) /i/



(c) /u/



(d) /o/

図 11. /e/ 音から生成した音声の周波数スペクトル（十代女性）

一方，以上と同様の手順によって /e/ 以外の母音から他の母音を生成することも試みたが，/a/，/i/，/u/，/o/ からは /e/ を生成することができなかった．また /i/ から生成した他母音は特に不明瞭な音声となり，明らかに判定困難な結果となった．

5．考　察

以下，前節で述べた実験結果について考察していく．本研究で用いた音声の加工方法は，PC 上のソフトウェアイコライザで各母音の周波数スペクトルの特徴に合わせて一部周波数帯域の振幅を減衰させているだけであるため，より多くの周波数成分を含む /e/ を他の母音から生成することが困難であったと考えられる．また，低周波域～3000Hz 近傍までの周波数成分が少ない /i/ については，必要帯域の振幅を減衰させるだけではスペクトル全体に対する周波数毎の相対的な変化を大きくできないために，明確に各母音の特徴を再現することが困難であったと考えられる．

この問題を解決するための一手法として，例えば，まずエンハンサで倍音成分を強調し十分な振幅を得たのちに，イコライザで母音毎に必要な周波数帯域の振幅を下げることにより，母音毎の周波数スペクトルの特徴をより明確にできる可能性がある．しかし一方で，加工回数を重ねるにしたがって声質の特徴を決める周波数成分も同時に損なってしまうことから，元の音声が持っている声色を損なう可能性が高くなることも考えられる．

今回のイコライザによる音声の加工では，それぞれのスペクトルの概形の特徴を際立たせるため，必要な帯域の倍音以外は振幅を－120[dB] まで下げ，全く聞こえない状態にした．一方で，基音，ならびに 4000Hz 以上の倍音は，母音判別に対する影響が少なく，また音質（元の声色）を維持するために重要な役割を持っていると判断し，なるべく振幅を下げずに残した．（基音のみを残した音を作成して聞いてみると，人の声とは認識しがたく正弦波のような電子音に近い音質になる．このことから，基音は母音判別に影響が少ないと判断した．）また，/i/ のような高周波成分が周波数スペクトルの多くを占める母音では，基音まで削ってしまうと「キーン」と感じられる不快な音になってしまうこともわかった．

6．まとめ

母音を決定付ける要素として周波数スペクトルの概形に着目し，一つの母音から他の４母音を生成することを試みた．その結果，/e/ 音の周波数スペクトルを他の各母音の特徴に合わせ加工することによって，残りの４母音を生成できる可能性があるという結果を得た．

本来，周波数スペクトルには声質に関する要素と母音に関する要素が混在しており，それらを分解し，解析を行う手順も考えられる．しかし，今回実際に生成した母音を評価した範囲では，それほど大きな声質の劣化があったわけではなく，元の声色も再現できていると考えられるので，よほどの精度が求められない限り，この点を無視しても実用性はあると考えられる．

今後の課題としては，さらに多くの母音サンプルを採取して実験を行い，母音変換フィルタのパラメータをより緻密に定めることによって，変換精度の向上を図りたい．また，この方法で作成した音声が正しく認識できるかどうかの普遍性についても，より多くの被験者の協力を得て確かめる必要があると考えている．

参考文献および参考ＵＲＬ

1）伊福部 達，音声タイプライタの設計―単音節音声認識の基礎と Z80 による製作例，ＣＱ出版，1983.

2）声と音の技術　声の種類と発生の仕組み（沖電気工業株式会社）http://www.oki.com/jp/rd/ss/speech.html（accessed 12 June 2015）

3）S.Rosen, P.Howell（著），荒井隆行，菅原勉（監訳），今富摂子他（訳），音声・聴覚のための信号とシステム，海文堂，1998.

4）重音テト「連続音」＋「単独音」ライブラリー http://kasaneteto.jp/ongen_down.html（accessed 12 June 2015）

5）林晃平，荒川正和，丸山晃生，“周波数スペクトルの概形に注目した母音変換”，北陸地区学生による研究発表会，F-82，2012.